

WHAT IS CLAIMED IS:

1. A method for identifying a unique electronic mail message in a plurality of electronic email messages extracted from an electronic mail messaging system, the method comprising:
 - retrieving a message from a mailbox on the electronic mail messaging system, the message including a plurality of message properties;
 - computing a message tag from at least a portion of the plurality of message properties;
 - reviewing a list of message tags stored in an index file; and
 - determining whether the message is unique based upon whether the message tag is found in an index file.
2. The method of claim 1, wherein the message tag is computed by concatenating at least two properties selected from the plurality of message properties.
3. The method of claim 2, wherein the message tag is further computed by applying a hash algorithm to the message tag to form a uniform string, wherein the uniform string has a pre-determined length.
4. The method of claim 3, wherein the hash algorithm is an MD5 hash algorithm.
5. The method of claim 1, wherein the plurality of message properties includes a sender's name and a sender's submission time, and wherein the message tag is computed by concatenating the sender's name to the sender's submission time.
6. The method of claim 1, wherein the plurality of message properties includes a sender's name, a sender's submission time and a subject, and wherein the message tag is computed by concatenating the sender's name and the subject to the sender's submission time.
7. The method of claim 1, wherein the index file is stored in a relational database system.
8. A method for archiving a plurality of electronic mail messages in a system external to an electronic mail messaging system, the method comprising:

reading a first message from a first mailbox on the electronic mail messaging system, the first message including at least a first sender's name and at least a first sender's submission time;

computing a first message tag from the first sender's name and the first sender's submission time;

storing the first message in a message archive and storing the first message tag in an index file associated with the message archive;

reading a second message from a second mailbox on the electronic mail messaging system, the second message including at least a second sender's name and at least a second sender's submission time;

computing a second message tag from the second sender's name and the second sender's submission time;

comparing the second message tag with the first message tag; and

storing the second message in the message archive and storing the second message tag in the index file if the first and second message tags are not the same.

9. The method of claim 8, wherein the first message tag is computed by concatenating the first sender's name and the first sender's submission time to form a first message string and wherein the second message tag is computed by concatenating the second sender's name and the second sender's submission time to form a second message string.

10. The method of claim 9, wherein the first message tag is further computed by applying a hash algorithm to the first message string to form a first uniform string, wherein the first uniform string has a pre-determined length, and wherein the second message tag is further computed by applying the hash algorithm to the second message string to form a second uniform string, wherein the second uniform string has the pre-determined length.

11. The method of claim 10, wherein the hash algorithm is an MD5 hash algorithm.

12. The method of claim 8, wherein the first mailbox and the second mailbox are different mailboxes on the electronic mail messaging system.
13. The method of claim 8, wherein the index file is stored in a relational database system.
14. The method of claim 8, wherein the message archive is a relational database system.
15. A system for identifying a unique electronic mail message, wherein the system is external to an electronic mail messaging system, the system comprising:
 - means for reading an electronic mail message from a mailbox on the electronic mail messaging system, the electronic mail message including a plurality of message properties;
 - means for computing a message tag from a least two properties selected from the plurality of message properties;
 - means for comparing the message tag with a list of message tags stored in an index file; and
 - means for determining that the message is unique if the message tag is not in the index file.
16. The system of claim 15, wherein the at least two properties comprise a sender's name and a sender's submission time.
17. The system of claim 15, wherein the message tag is computed by concatenating the at least two properties to form a first message string.
18. The system of claim 17, wherein the message tag is further computed by applying a hash algorithm to the message string to form a uniform string, wherein the uniform string has a pre-determined length.
19. The system of claim 18, wherein the hash algorithm is an MD5 hash algorithm.
20. The system of claim 15, wherein the index file is stored in a relational database system.
21. A system for identifying a unique electronic mail message, wherein the system is external to an electronic mail messaging system, the system comprising:

a uniqueness checker in communication with the electronic mail messaging system;
and

an index file comprising a plurality of pre-determined message tags,
wherein the uniqueness checker is configured to read a message from the electronic mail messaging system, wherein the message includes a plurality of properties associated with the message,

wherein the uniqueness checker computes a message tag for the message using at least two of the properties, and compares the computed message tag with the index file,

wherein if the computed message tag matches an entry in the index file, the uniqueness checker determines that the message is not unique, otherwise, if the computed message tag does not match an entry in the index file, the computed message tag is added to the index file.

22. The system of claim 21, wherein the message tag is computed by concatenating the at least two properties to form a message string.

23. The system of claim 22, wherein the message tag is further computed by applying a hash algorithm to the message string to form a uniform string, wherein the uniform string has a pre-determined length

24. The system of claim 23, wherein the hash algorithm is an MD5 hash algorithm.

25. The system of claim 21, wherein the uniqueness checker reads the message from a mailbox on the electronic mail messaging system.

26. The system of claim 21, wherein the plurality of properties comprises a sender's name and a sender's submission time.

27. The system of claim 26, wherein the plurality of properties further comprises a subject string, and wherein the message tag is computed by concatenating the sender's name, the sender's submission time, and the subject string to form a message string.

28. The system of claim 27, wherein the message tag is further computed by applying a hash algorithm to the message string to form a uniform string, wherein the uniform string has a pre-determined length.

29. The system of claim 15, wherein the index file is stored in a relational database system.

30. A system for archiving a plurality of electronic mail messages, wherein the system is external to an electronic mail messaging system, the system comprising:

means for reading a first message from a first mailbox on the electronic mail messaging system, the first message including at least a first sender's name and at least a first sender's submission time;

means for computing a first message tag from the first sender's name and the first sender's submission time;

means for storing the first message in a message archive and storing the first message tag in an index file associated with the message archive;

means for reading a second message from a second mailbox on the electronic mail messaging system, the second message including at least a second sender's name and at least a second sender's submission time;

means for computing a second message tag from the second sender's name and the second sender's submission time;

means for comparing the second message tag with the first message tag; and

means for storing the second message in the message archive and storing the second message tag in the index file if the first and second message tags are not the same.

31. The system of claim 30, wherein the first message tag is computed by concatenating the first sender's name and the first sender's submission time to form a first message string and wherein the second message tag is computed by concatenating the second sender's name and the second sender's submission time to form a second message string.

32. The system of claim 31, wherein the first message tag is further computed by applying a hash algorithm to the first message string to form a first uniform string, wherein the first uniform string has a pre-determined length, and wherein the second message tag is further computed by applying the hash algorithm to the second message string to form a second uniform string, wherein the second uniform string has the pre-determined length.
33. The system of claim 32, wherein the hash algorithm is an MD5 hash algorithm.
34. The system of claim 30, wherein the first message further comprises a first subject string and the second message further comprises a second subject string, and wherein the first message tag is computed by concatenating the first sender's name, the first sender's submission time, and the first subject string to form a first message string, and wherein the second message tag is computed by concatenating the second sender's name, the second sender's submission time and the second subject string to form a second message string.
35. The system of claim 30, wherein the index file is stored in a relational database system.
36. The system of claim 30, wherein the message archive is a relational database system.
37. A system for externally archiving a plurality of electronic mail messages selected from an electronic mail messaging system, the system comprising:
- an archive server in communication with the electronic mail messaging system;
 - a uniqueness checker in communication with the archive server; and
 - an archive message store in communication with the archive server,
- wherein when the archive server reads a message from the electronic mail messaging system, a plurality of properties associated with the message are sent from the archive server to the uniqueness checker,
- wherein the uniqueness checker computes a message tag for the message using at least two of the properties, and compares the computed message tag with an index file,

wherein if the computed message tag matches an entry in the index file, the uniqueness checker indicates to the archive server that the message is not unique, otherwise, if the computed message tag does not match an entry in the index file, the computed message tag is added to the index file,

wherein if the message is unique, the archive server stores the message in the archive message store.

38. The system of claim 37, wherein the message tag is computed by concatenating the at least two properties to form a message string.

39. The system of claim 38, wherein the message tag is further computed by applying a hash algorithm to the message string to form a uniform string, wherein the uniform string has a pre-determined length

40. The system of claim 39, wherein the hash algorithm is an MD5 hash algorithm.

41. The system of claim 37, wherein the archive server reads the message from a mailbox on the electronic mail messaging system.

42. The system of claim 41, wherein the plurality of properties comprises a sender's name and a sender's submission time.

43. The system of claim 42, wherein the plurality of properties further comprises a subject string, and wherein the message tag is computed by concatenating the sender's name, the sender's submission time, and the subject string to form a message string.

44. The system of claim 43, wherein the message tag is further computed by applying a hash algorithm to the message string to form a uniform string, wherein the uniform string has a pre-determined length.